Visualization of Multivariate Physiological Data for Cardiorespiratory Fitness Assessment through ECG (R-Peak) Analysis

J. E. Muñoz, S. Bermudez i Badia, E. Rubio, M. S. Cameirão

Abstract— The recent rise and popularization of wearable and ubiquitous fitness sensors has increased our ability to generate large amounts of multivariate data for cardiorespiratory fitness (CRF) assessment. Consequently, there is a need to find new methods to visualize and interpret CRF data without overwhelming users. Current visualizations of CRF data are mainly tabular or in the form of stacked univariate plots. Moreover, normative data differs significantly between gender, age and activity, making data interpretation yet more challenging. Here we present a CRF assessment tool based on radar plots that provides a way to represent multivariate cardiorespiratory data from electrocardiographic (ECG) signals within its normative context. To that end, 5 parameters are extracted from raw ECG data using R-peak information: mean HR, SDNN, RMSSD, HRVI and the maximal oxygen uptake, VO_{2max}. Our tool processes ECG data and produces a visualization of the data in a way that it is easy to compare between the performance of the user and normative data. This type of representation can assist both health professionals and non-expert users in the interpretation of CRF data.

I. Introduction

Cardiorespiratory fitness (CRF) or cardiorespiratory endurance relates to the ability of the human circulatory, musculoskeletal and respiratory systems to supply oxygen during exercise [1]. It is normally considered a crucial parameter to assess fitness in order to prescribe exercises, pharmacological treatments and rehabilitation therapies. One important aspect of CRF is that it directly relates to an individual's functional status. Some of the parameters for the assessment of CRF can be derived from physiological measurements such as an electrocardiogram (ECG) which can be recorded during exercise. However, despite the relevance of the prognostic information provided by ECG based assessments of fitness, they are rarely used in primary care settings [2], possibly due to the cost and non-portability characteristics of the devices and the problems related with signal recording during exercise.

The recent rise in the use of wearable and ubiquitous fitness sensors has allowed the increase of behavioral logs

J.E. Muñoz, S. Bermudez i Badia and M. Cameirão are with the Madeira ITI, Universidade da Madeira, Funchal, Portugal (e-mails: {john.cardona, sergi.bermudez, monica.cameirao}@miti.org). E. Rubio is with Department of Physical Education and Sport, Universidade da Madeira, Funchal Portugal (e-mail: erubiog@uma.pt).

which can be defined as digital portraits of behavioral patterns recorded via multiple physiological sensors [3]. The use of this low-cost technology has made it extremely easy to collect and store data, and the companion web applications let you access them whenever you want. These behavioral logs may include cardiac information, location, and/or movement and performed activities. Unfortunately, in the particular case of CRF, the existing data visualization tools are still insufficient for non-expert users and there is a need for new methods to visualize and interpret CRF data.

Users of these novel tools to measure and register physiological signals want to understand what their data says about them. This created new approaches to facilitate the interpretation of large amounts of data. One of the best ways to explore and try to understand large datasets is by developing novel information visualization tools to reinforce the understanding of abstract data [4]. In this case, the challenge is not necessarily the use of computational methods to display large amounts of data in nice looking graphs, but rather to augment human perception and the understanding of data through their visual representation. For this reason, visual representation of data plays an important role in the confluence between the availability of complex multivariate health data and the rise and quick adoption of health tracker technology [5]. However, this phenomenon is not unique to health trackers. In the medical context, as our ability to generate and access abundant clinical data is growing at an exciting pace, it also becomes imperative to develop new methods to organize and visualize information without overwhelming users. Health professionals are faced with an information overload of specialized medical data and the way data are presented affects how such information is interpreted [6]. Consequently, the following questions arise: how can we improve the understanding and interpretation of physiological signals usually represented in stacked univariate plots? How do these signals relate to the specific condition that is being assessed? How can non-experts users interpret their own recorded signals?

The central goal of this work was to design and develop an interactive information visualization tool to represent the data from a wearable electrocardiography sensor, by extracting a specific set of parameters to assess cardiorespiratory fitness, and interpreting it through pre-established normative data [7]. The main design constrains of this tool are related with i) representation of the data in a way that it is easy to compare between the current state of the user and normative data; ii) representation that is useful for both non-expert users and health professionals; and iii) the visualization has to be

generated with minimum computer resources within an acceptable period of time.

II. METHODS: CARDIORESPIRATORY FITNESS ASSESSMENT FROM ELECTROCARDIOGRAPHY DATA

ECG registers by means of surface electrodes the electrical activity generated by the dynamics of the myocardial fibers of the heart. The advantages of this bioelectrical measurement methodology are its low cost, real time feedback, low invasiveness and easy implementation. The ECG is widely used to diagnose some types of abnormal heart function, including arrhythmias and conduction disturbances [8]. Through ECG measurements, important parameters for the CRF assessment can be extracted. In our particular case, 5 parameters derived of the ECG signal were chosen. The different formulas to compute the parameters were chosen considering existent bibliography for healthy adults.

The detection of the QRS complex is one of the most important steps carried out in the ECG signal analysis. The detection of R peaks within the QRS complex is easier than the other ECG segments due to its morphology because of its structural form and high amplitude. From it, the RR interval (RRI) – the time duration between two successive R peaks – can also be computed [9]. For a healthy adults, the average number of the heart beats per minute is about 60-80, which corresponds to a mean value of RRI about 600-750 ms [8]. The mean HR is a simple time-domain variable that can be easily derived from ECG signals recorded in normal conditions and by low cost personal trackers. RR intervals can be converted to HR by the following equivalent formula [8]:

$$HR \ in \ beats/\min = 1000/RR \ (in \ ms) \times 60 \qquad (1)$$

Heart Rate Variability (HRV) can be measured through variables such as the standard deviation of the RR interval (SDNN, sometimes also called SDRR) and the square root of the mean squared differences of successive RR intervals (RMSSD). SDNN represents the extent of heart rate variability and normally is used to measure temporal changes in the heart behavior for 24-hour period [9].

$$SDNN = \sqrt{\frac{1}{N-1}\sum_{n=2}^{N}[I(n) - \bar{I}]^2} \qquad (2)$$

where I is the RR interval and N the total number of all RR intervals in the segment. RMSSD represents the difference between consecutive R waves (eq. 3). Thus, when RMSSD decreases it means that the person's heart beat becomes quicker.

$$RMSSD = \sqrt{\frac{1}{N-2}\sum_{n=3}^{N}[I(n) - I(n-1)]^2} \qquad (3)$$

There is a strong dependence of HRV parameters with age, which means that the values require either to be adjusted for age or that age-related references need to be considered [10]. Because of these dependencies, extensive research has been devoted to the consolidation of normative data for SDNN and RMSSD for different age ranges [11].

Geometric methods in HRV analysis include parameters derived from RR histograms. The shape of the histograms has been associated with mood factors [11]. The HVR triangular index (HRVI) is computed as a triangle with its baseline width conformed by the amount of RR interval variability and its height corresponding to the peak of the histogram count of all RR intervals [9].

$$HRVI = \frac{\sum_{i=1}^{Nb} b(t_i)}{max_i b(t_i)} \qquad (4)$$

where N_b is the number of bins and the number of intervals in the ith bin centered at t_i is represented as $b(t_i)$.

Finally, another important factor to evaluate functional cardiorespiratory capacity is VO_{2max}. VO_{2max} is the maximum oxygen uptake capacity and it is a reliable measurement that reflects the functional capacity of the cardiorespiratory system. This measure represents the ability of the body's circulatory and respiratory systems to deliver oxygen to the muscles during dynamic exercising involving a large muscle mass [7]. The estimation of the VO_{2max} is performed using the ratio between the heart rate values at rest (HR_{rest}) and maximal exercise (HR_{max}). The value of HR_{rest} is defined as the lowest value of any 1-min average during a 5-min sampling period, and the HR_{max} can be approximated using the highest 5-s average during different exercise tests such as a treadmill test. Additionally, the HR_{max} can be estimated using several age-related formulas [12]. Hence, VO_{2max} has been used extensively as a simple and useful tool to estimate cardiorespiratory capacity [13].

$$VO_{2max} = (15.0 \ ml \cdot min^{-1} \cdot Kg^{-1}) \times \frac{HR_{max}}{HR_{rest}} \qquad (5)$$

VO_{2max} is used to classify the cardiorespiratory fitness level (poor, fair, good, excellent or superior) using age/gender related normative values [7]. Thus, many performance tests have been devised to predict VO_2max directly from ECG measurements [7].

III. METHODS: MULTIVARIATE PHYSIOLOGICAL DATA VISUALIZATION

All the above ECG parameters play an important role in the CRF assessment. However, the current methods to visualize multivariate physiological data focus on visualizing independent or stacked temporal windows that represent individual parameters. This often results in confusing or overloaded graphs, complicating the assessment by the specialist and limiting the understanding of final users about their cardiorespiratory status. Nevertheless, there are other multivariate temporal visualization techniques designed to facilitate data interpretation. One method that has recently become popular is the use of radial plotting to visualize multivariate physiological data [14]. StartPlots or RadarPlots are graphical methods for displaying multivariate data in the
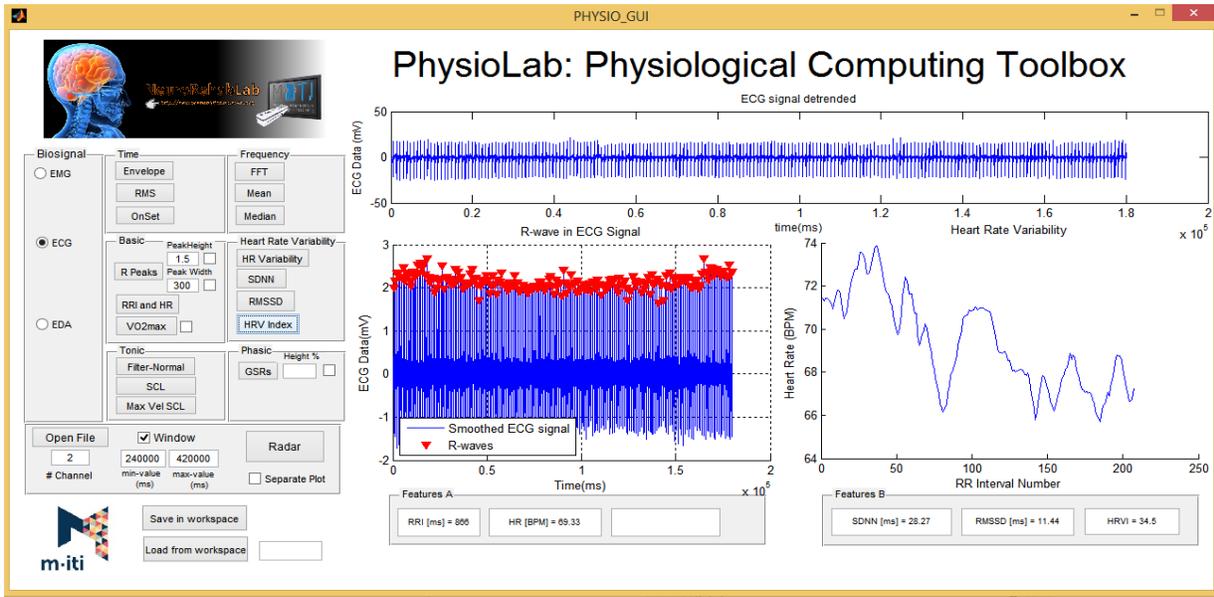
Figure 1. The PhysioLab, an interactive information visualization toolbox developed to assist in the assessment of cardiorespiratory fitness from wearable electrocardiographic data.

form of a 2D chart of three or more quantitative variables represented on axes starting from the same point [4]. Each multivariate observation can be seen as a data point in an n-dimensional vector space:

- Arrange $\mathbb{N}$ axes on a circle in $\mathbb{R}^N$ and $3 \leq \mathbb{N} \leq \mathbb{N}_{max}$
- Map coordinate vectors $P \in \mathbb{R}^N$ from $\mathbb{R}^N \rightarrow \mathbb{R}^2$
- $P = \{p_1, p_2, ..., p_N\} \in \mathbb{R}^N$ where each p_i represents a different attribute with a different physical unit.
- Each axis represents one attribute of data. Each data record or data point P is visualized by a line along the data points.
- A line is perceived better than just points on the axes.

In these plots, several axes can be drawn, one for each variable, starting from the middle and equally spaced in a circle. The center represents the minimum value for each variable, and the ends represent the corresponding maximum values.

Past work has shown the usefulness radar plots in multivariate physiological data representation. Sebastian et al. created a multi-signal visualization tool for traumatic brain injury monitoring [15]. Physiological data related with intracranial pressure, arterial pressure, cerebral perfusion pressure, heart rate and cerebral tissue oxygenation are plotted as polygrams (or radars) as new visualization dashboards. Another more recent work by Holzinger et al. shows the design and development of an end-user centered mobile software for interactive visualization of physiological data to support the prevention of stress and burnout syndromes [16]. ECG and EDA signals are plotted in star plots, the values of each feature being normalized to provide a better comparison between plots. The presentation of health care data in star plots for clinical studies has been shown to be useful because

it allows the easy assessment of [14]: i) changes over time in multiple variables, in a single individual or in different groups, ii) multiple- treatment group differences on multiple-outcome measures, iii) difference between disease conditions on multiple variables.

In our solution, we considered the following normative values for: HR [7], SDNN, RMSSD, HRVI [11] and VO_{2max} [7]. In order to personalize the visualization to each user and facilitate data interpretation, each radar plot is generated taking into account 3 important aspects: age, gender and activity, which normally define the normative reference values. Further, 3 fitness profiles were defined for cardiorespiratory performance: high (Excellent-Superior), normal (Good), and low (Poor-Fair). For these, all axes are divided in 3 equally sized frames considering the values of the 3 profiles for each parameter. Prior to rendering the visualization, a uniform scale is used on every axes for all parameters and for each of the 3 profile divisions. This approach provides an elegant solution to the "filled in" problem of radar plots, in where the area and shapes of the formed representations change dramatically with data ranges, non-linearity of data, and the order of each parameter in the graph axes.

IV. RESULTS: VISUALIZATION TOOL IMPLEMENTATION

We developed the PhysioLab, an interactive information visualization toolbox to represent ECG parameters for CRF assessment using radar plots (Figure 1). Following the above described methodology, we implemented 3 stages prior to obtaining the final visual representation: ECG signal processing, feature extraction, and data normalization according to the corresponding normative data. The PhysioLab toolbox and the processing of the ECG signals are implemented in Matlab (MathWorks Inc., Massachusetts,

US). PhysioLab uses different signal pre-processing methods to reduce low frequency artifacts caused by respiration and user movements. To stabilize ECG baseline wander, which is characterized for baseline oscillations at a very-low frequency drifting between 0.15 Hz and 0.3 Hz, some denoising techniques are used. First, the signal is smoothed by a fifth-order Savitzky-Golay FIR filter. Then, a low order polynomial is fitted to the raw data and is used to detrend the ECG signal. At this stage, a simple threshold rule based on the peak morphology is applied to detect the R-peaks. PhysioLab presents the original signal and the detrended one with the R-peaks detection in separate windows (Figure 1). All ECG parameters are displayed within the user interface of the PhysioLab toolbox and stored to produce the Radar Plot. Figure 1 shows a previsualization of the toolbox after ECG signal processing and feature extraction. Once all CRF parameters are extracted from the raw ECG, a radar plot is generated (Figure 2).
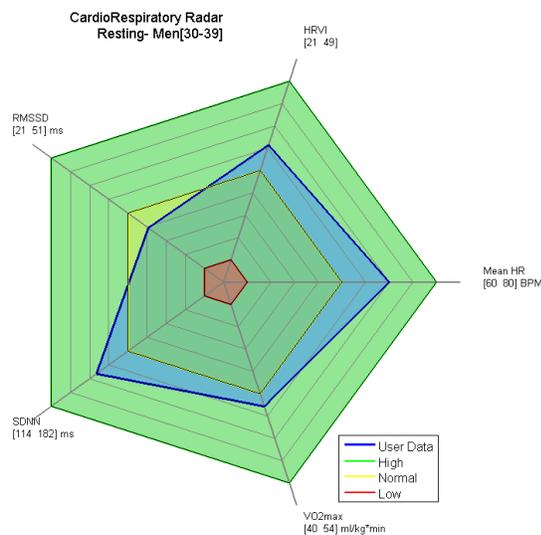

Figure 2. An example cardiorespiratory radar plot generated using MeanHR, VO$_{2max}$, SDNN, RMSSD and HRVI parameters.

Performance profiles derived from available normative data are generated and color-coded facilitating the interpretation and comparison of user performance (blue line) with the CRF levels (high-green, average-yellow, low-red). Figure 2 shows the cardiorespiratory radar for a five minutes resting ECG recordings of a 32 years old woman. The final radar plot allows for a quick visualization of the cardiorespiratory fitness state of the user for the activity proposed. In this case, almost all the parameters are in the high profile of CRF indicating an excellent-superior classification of the user performance. Only the RMSSD parameter reported lower values than the high profile but still within a "safe" average zone (see yellow zone in figure 2).

CONCLUSIONS

Here we presented a tool for the assessment of cardiorespiratory fitness using ECG data, and a novel approach to visualize multivariate physiological data. The solution proposed is a Matlab toolbox that extracts five cardiorespiratory parameters from pre-processed ECG signals. Our visualization improves on the classic radar plot concept and applies it in the CRF assessment domain. Our visualization method normalizes data in each axis of the plot and establishes color-coded zones to represent the different fitness profiles. These are then compared with the user's current performance in a particular activity. Such strategy allows expert users to quickly determine the classification of a person's activity performance in terms of functional cardiorespiratory capacity, which is defined by tabulated normative data. In addition, end-users can easily understand their CRF profile without the necessity of interpreting data or remembering normative ranges of values. Finally, using the Physiolab toolbox, the generation of the radar plot is generated instantly from the ECG features previously extracted and from the normative data defined by our literature review. Future work should include interactivity in order to navigate to past user records.

REFERENCES

[1] A. C. o. S. Medicine, *ACSM's health-related physical fitness assessment manual*: Lippincott Williams & Wilkins, 2013.
[2] P. P. Toth and C. P. Cannon, *Comprehensive cardiovascular medicine in the primary care setting*: Springer Science & Business Media, 2010.
[3] S. Dumais, R. Jeffries, D. M. Russell, D. Tang, and J. Teevan, "Understanding User Behavior Through Log Data and Analysis," in *Ways of Knowing in HCI*, ed: Springer, 2014, pp. 349-372.
[4] A. Holzinger, "Lecture 9 Interactive Information Visualization and Visual Analytics," in *Biomedical Informatics*, ed: Springer, 2014, pp. 379-420.
[5] P. Taylor, *From patient data to medical knowledge: the principles and practice of health informatics*: John Wiley & Sons, 2008.
[6] A. A. Bui and W. Hsu, "Medical Data Visualization: Toward Integrated Clinical Workstations," in *Medical Imaging Informatics*, ed: Springer, 2010, pp. 139-193.
[7] V. H. Heyward and A. Gibson, *Advanced Fitness Assessment and Exercise Prescription 7th Edition*: Human Kinetics, 2014.
[8] A. L. Goldberger, *Clinical electrocardiography: a simplified approach*: Elsevier Health Sciences, 2012.
[9] G. Ernst, *Heart rate variability*: Springer, 2014.
[10] A. Voss, A. Heitmann, R. Schroeder, A. Peters, and S. Perz, "Short-term heart rate variability—age dependence in healthy subjects," *Physiological measurement,* vol. 33, p. 1289, 2012.
[11] P. W. Macfarlane, A. van Oosterom, O. Pahlm, P. Kligfield, M. Janse, and J. Camm, *Comprehensive electrocardiology*: Springer Science & Business Media, 2010.
[12] H. Tanaka, K. D. Monahan, and D. R. Seals, "Age-predicted maximal heart rate revisited," *Journal of the American College of Cardiology,* vol. 37, pp. 153-156, 2001.
[13] N. Uth, H. Sørensen, K. Overgaard, and P. K. Pedersen, "Estimation of VO2max from the ratio between HRmax and HRrest–the heart rate ratio method," *European journal of applied physiology,* vol. 91, pp. 111-115, 2004.
[14] M. J. Saary, "Radar plots: a useful way for presenting multivariate health care data," *Journal of clinical epidemiology,* vol. 61, pp. 311-317, 2008.
[15] K. Sebastian, V. Sari, L. Y. Loy, F. Zhang, Z. Zhang, and M. Feng, "Multi-signal visualization of physiology (MVP): a novel visualization dashboard for physiological monitoring of traumatic brain injury patients," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 2000-2003.
[16] A. Holzinger, M. Bruschi, and W. Eder, "On Interactive Data Visualization of Physiological Low-Cost-Sensor Data with Focus on Mental Stress," in *Availability, Reliability, and Security in Information Systems and HCI*, ed: Springer, 2013, pp. 469-480.